The early emergence of guilt-motivated prosocial behavior

Supplementary materials

The early emergence of guilt-motivated prosocial behavior

Amrisha Vaish,[1,2] Malinda Carpenter,[1,3] and Michael Tomasello[1]

[1]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[2]University of Virginia

[3]University of St Andrews, St Andrews, Scotland

In press in *Child Development*.

Detailed results

Experimenters

All 3-year-old children were tested by the same two female experimenters, but the experimenter who played E1's role for the 3-year-old children unexpectedly departed after testing 8 2-year-old children. A new E1 thus tested the remaining 2-year-old children. Comparisons of behavior between 2-year-old children tested by the two different experimenters revealed no significant differences on any of the measures, thus ruling out an experimenter effect.

E1's emotional displays

The sad and neutral expressions displayed by E1 during E1's return and Questions phases were based on the sad and neutral expressed displayed by experimenters in prior work with children of similar ages (Vaish, Carpenter, & Tomasello, 2009, 2010) To ensure that naïve individuals could recognize E1's emotional expressions as matching the intended sad or neutral displays, we conducted a manipulation check on approximately 10% of our sample (n = 13). These 13 children were randomly selected with the constraints that about half should be from each age group and about half in the Harm conditions. Thus, of the 13 children, 7 were 2-year-olds (4 in the Harm conditions) and 6 were 3-year-olds (3 in the Harm conditions). Additionally, of the 4 2-year-olds who were in the Harm conditions, 2 children were tested by the first E1 and 2 were tested by the second E1 (see above).

We presented the E1's return and Questions phases of these 13 videos without sound (as the content of E1's speech would give away whether there was harm or no harm, thus providing a cue as to which emotion she was intended to display) and in a random order to an independent adult coder (who was blind to condition). The coder judged whether E1 was displaying a happy, neutral, sad, or angry emotion. This coder correctly identified E1's facial expressions as sad or

neutral, appropriately for each condition, in all 13 videos, indicating that E1's emotional expressions were consistent and matched the intended displays.

Gender

Preliminary analyses of both primary and secondary measures revealed only three effects of gender. First, in the analysis of children's looks to E2 during the 15 s following the mishap, gender interacted with cause, outcome, and age, and was thus retained in that analysis (see below for details). In addition, during E1's return and Questions phases (combined), girls looked significantly more to E1 than boys, and on the first question ("What happened?"), 3-year-old girls were more likely than boys to make statements about the mishap (e.g., "Tower broke"). As there were no further effects of gender, this variable was not included in analyses except in the one mentioned above.

Guilt-relevant reparative behavior

15 s following accident

As guilt is an inherently social and interpersonal emotion (Baumeister, Stillwell, & Heatherton, 1994), our primary focus in this study was on analyzing the reparative and prosocial behaviors that children showed once E1 returned to the scene and was available for children to express guilt towards (or not), i.e., beginning with E1's return. However, once E1 returned, the four conditions were not identical, since E1 expressed sadness when she was harmed and had a slightly different verbal response to the mishap in the Harm versus No Harm conditions. Reviewers thus recommended that we additionally examine guilt-relevant reparative behavior in the brief (approx. 15 s) period between the accident and E1's return – when the conditions were identical (since E1 had not yet responded to the mishap). We thus carried out a secondary analysis of children's guilt-relevant reparative behavior during this initial period.

We assigned children a score according to whether or not they produced either physical or verbal reparative behavior during the 15 s immediately following the accident (prior to E1's return). Children received 1 point if they attempted to repair the tower and 1 point if they expressed guilt in their speech. Each child thus received a score from 0 to 2. As a large number of children received scores of 0 (see Table S2), the data were positively skewed and were thus analyzed using non-parametric tests (applying the Monte Carlo permutation method).

For each age group, Kruskal-Wallis tests were conducted with the guilt-relevant reparative behavior score as the dependent variable and condition as the independent variable. These analyses indicated significant effects of condition in both age groups (2-year-old: $H(3) = 9.20$, $p = .021$; 3-year-olds: $H(3) = 9.13$, $p = .048$). However, these were not due to a guilt-specific effect. Among 2-year-olds, pairwise comparisons using Mann-Whitney U tests indicated that children in the Child-No Harm condition scored significantly higher than children in the E2-Harm condition (Child-No Harm: $M = .38$, $SD = .50$; E2-Harm: $M = .00$; $U = 75$, $p = .017$). None of the other pairwise comparisons revealed significant differences, all ps $> .076$. In particular, although the scores were highest in Child-No Harm, they were only slightly lower in Child-Harm. Among 3-year-olds as well, although scores were highest in Child-No Harm, pairwise comparisons did not reveal any significant differences between conditions, all ps $> .098$.

Thus, in the 15 s immediately following the accident, when E1 had not yet returned to the scene, we did not find evidence of a guilt-specific effect in children's guilt-relevant reparative behavior in either age group. At first glance, this might seem to contradict our finding that 3-year-olds showed guilt-specific reparative behavior in the E1's return and Questions phases. However, we do not think this is a surprising effect: Given the inherently social and interpersonal nature of guilt (Baumeister et al., 1994), it makes sense for it to be expressed (verbally and/or as

reparative behavior) primarily in the presence of the victim and to be expressed less (if at all) when the victim is absent. Certainly, this need not mean that children do not feel guilty if the victim is absent, but rather that their guilt is less likely to manifest itself in verbal or non-verbal reparative behavior in the victim's absence. This tentative proposal is based on our understanding of the functions of guilt and the suggestive evidence presented above, but it should be examined systematically in future research.

E1's return and Questions phases (combined): Verbal versus non-verbal behavior

For these analyses, we gave children a separate score for verbal and non-verbal reparative behavior in the first two phases. Thus, in both the E1's return and the Questions phases, children received 1 point if they attempted to repair the tower (for a maximum of 2 points on the non-verbal measure), and similarly, 1 point if they expressed guilt in their speech (for a maximum of 2 points on the verbal measure). Table S3 provides detailed descriptive data.

For each age group, two Kruskall-Wallis tests (applying the Monte Carlo permutation method) were conducted – one with the verbal and one with the non-verbal score as the dependent variable, and condition as the independent variable. The results paralleled those from the analyses of combined verbal and non-verbal scores. Among 3-year-olds, there was a significant effect of condition for both measures (verbal: $H(3) = 13.09$, $p = .004$; non-verbal: $H(3) = 12.69$, $p = .007$), whereas among 2-year-olds, there was no significant effect of condition on either measure (verbal: $H(3) = 2.88$, $p = 1.00$; non-verbal: $H(3) = 4.23$, $p = .266$). It thus does not seem to be the case that our combined verbal and non-verbal score simply washed out any guilt-specific effects in 2-year-olds' non-verbal reparative behaviors.

Stickers

Shared half or more versus less than half

For each age group, we conducted 2 (outcome) x 2 (cause) logistic regressions using whether children shared half (i.e., 2) or more versus less than half of their stickers with E1 as the dependent measure. Among 3-year-olds, the full model revealed only a significant main effect of outcome, $z = 2.14$, $p = .033$. Children at this age were more likely to share half or more of their stickers with E1 if E1 had previously been harmed (10 of 32 children) than if she had not been harmed (4 of 32 children), thus again indicating an effect of sympathy but not of guilt (see Table S5). Among 2-year-olds, the full logistic regression model was not significantly different from the null model, $p = .788$, and was thus not analyzed further.

Secondary analyses

Recent work on sharing suggests that young children tend to hoard resources such as stickers and to not share them equally, even when they know that they should share (e.g., Smith, Blake, & Harris, 2013). We explored whether such a pattern emerged in the sharing behavior in our study as well. We first examined whether or not children shared at all (i.e., gave one or more stickers to E1). We began with a 2 (age) x 2 (outcome) x 2 (cause) logistic regression model. As including the three-way interaction, the outcome x cause interaction, and the age x cause interaction did not predict the data better than the reduced models without each of these terms, respectively (all ps > .104), these terms were removed from the final model. Including the age x outcome interaction did predict the data better than the reduced model without this interaction term ($p = .026$), so this two-way interaction was retained. The final model thus included the age x outcome interaction and all three main effects.

This final model revealed only a significant main effect of age, $z = 3.06$, $p = .002$ (the age x outcome interaction, when entered into the final model along with the three main effects, was at a non-significant trend, $p = .071$, and the main effects of cause and outcome were both also not significant, both ps > .79). The main effect of age was due to the fact that a greater proportion of 3-year-olds (55 of 64) than 2-year-olds (23 of 62) shared stickers with E1. Thus, whether or not children shared was not significantly impacted by guilt or sympathy, but the likelihood of sharing at all did increase between 2 and 3 years such that by 3 years, most children shared at least one sticker with E1.

Interestingly, however, one-sample *t*-tests (test value = 2 stickers, out of the maximum of 4) revealed that overall, both 2- and 3-year-olds shared significantly less than the equal split of 2 stickers (2-year-olds: $M = .58$, $SD = .95$, $t(61) = 11.76$, $p < .0005$; 3-year-olds: $M = 1.16$, $SD = .80$, $t(63) = 8.43$, $p < .0005$). This finding supports Smith et al.'s (2013) finding that preschool-aged children do not engage in equal sharing.

Looking

Here we provide detailed results of the proportion of time children spent looking to E1, E2, and the broken tower during the 15 s following the accident as well as during E1's return and Questions phases. For all analyses of looking, we began with a 2 (cause) x 2 (outcome) x 2 (age) x 2 (gender) ANOVA. If this revealed no significant effects of gender, we removed gender from the analysis and conducted a 2 (cause) x 2 (outcome) x 2 (age) ANOVA. If age was found to interact significantly with any other factors, we analyzed the age groups separately.

15 s following accident

Looks to E1. This ANOVA revealed a cause x outcome x age interaction, $F(1,118) = 5.73$, $p = .018$, $\eta_p^2 = .046$. Among 2-year-olds, there were no main effects or interactions, all ps

> .161. Among 3-year-olds, there was a significant cause x outcome interaction, $F(1,60) = 4.92$, $p = .030$, $\eta_p^2 = .076$: Three-year-olds looked to E1 a similar percentage of time when they had caused harm versus no harm (Child-Harm: $M = 10.38$, $SD = 10.81$; Child-No Harm: $M = 11.51$, $SD = 20.43$; $p = .85$), whereas they looked to E1 more when E2 had caused harm than when E2 had caused no harm (E2-Harm: $M = 15.31$, $SD = 13.09$; E2-No Harm; $M = 1.64$, $SD = 2.65$; $t(30) = 4.10$, $p < .0005$). Notably, children's looking to E1 was significantly or marginally significantly lower in E2-No Harm than in all other conditions ($p = .069$ compared to Child-Harm; $p = .041$ compared to Child-No Harm; $p = .005$ compared to E2-Harm), whereas the other three conditions did not differ significantly from one another (all ps > .30). These results suggest that 3-year-olds' looking to E1 during this initial phase was impacted by both cause and outcome, and was substantially reduced only if neither their own agency nor a harmful outcome were involved.

Looks to E2. For this measure, we obtained a cause x outcome x age x gender interaction, $F(1,110) = 6.05$, $p = .015$, $\eta_p^2 = .052$. Among 2-year-olds, there was only a main effect of cause, $F(1,54) = 4.38$, $p = .041$, $\eta_p^2 = .075$, such that 2-year-olds looked more to E2 if E2 had caused the accident ($M = 18.97$, $SD = 2.58$) than if they themselves had caused it ($M = 11.47$, $SD = 2.49$). Thus, immediately following the accident, 2-year-olds tracked who caused the accident.

Among 3-year-olds, there was a cause x outcome x gender interaction, $F(1,56) = 12.61$, $p = .001$, $\eta_p^2 = .184$. Analyzing girls and boys separately revealed no significant effects among girls (all ps > .104) but a significant cause x outcome interaction among boys, $F(1,28) = 13.64$, $p = .001$, $\eta_p^2 = .328$: In the harm conditions, boys spent a greater proportion of time looking to E2 if E2 had caused harm ($M = 22.46$, $SD = 10.85$) than if they had caused harm ($M = 10.10$, $SD =$

11.45; $t(14) = 2.22$, $p = .044$), whereas they spent a greater proportion of time looking to E2 if they themselves had caused no harm ($M = 19.27$, $SD = 14.20$) than if E2 had caused no harm ($M = 2.46$, $SD = 6.95$; $t(14) = 3.01$, $p = .013$). These results again suggest that in the 15 s immediately following the accident (when the conditions were identical), 3-year-olds were sensitive to both of the critical factors underlying guilt, namely, cause and outcome.

Looks to tower. For this measure, we found a cause x age interaction, $F(1,118) = 7.84$, $p = .006$, $\eta_p^2 = .062$. Among 2-year-olds, there was only a significant main effect of cause, $F(1,58) = 14.17$, $p < .0005$, $\eta_p^2 = .196$: These children spent a greater proportion of time looking to the broken tower if they had caused the mishap ($M = 46.54$, $SD = 25.67$) than if E2 had caused it ($M = 25.44$, $SD = 17.68$). This corroborates the conclusion that immediately following the accident, 2-year-olds differentiated between when they versus E2 had caused the accident. Among 3-year-olds, there was a significant cause x outcome interaction, $F(1,60) = 4.30$, $p = .042$, $\eta_p^2 = .067$. These children looked to the broken tower a similar percentage of time when they or E2 had caused a harmful outcome (Child-Harm: $M = 19.76$, $SD = 11.40$; E2-Harm: $M = 23.07$, $SD = 13.68$; $p = .464$), whereas they looked to the tower more if they had caused a non-harmful outcome than if E2 had done so (Child-No Harm: $M = 28.44$, $SD = 14.83$; E2-No Harm; $M = 18.65$, $SD = 10.06$; $t(30) = 2.19$, $p = .037$).

E1's return and Questions phases (combined)

Looks to E1. This analysis revealed only a main effect of outcome, $F(1,118) = 22.33$, $p < .0005$, $\eta_p^2 = .159$. Across ages, children looked significantly more to E1 if she had been harmed ($M = 35.54$, $SD = 16.14$) than not harmed ($M = 22.19$, $SD = 15.26$). Thus, once E1 returned to the scene, children at both ages distinguished between the harm and no harm conditions. This is particularly significant for the 2-year-olds who, prior to E1's return, only showed evidence of

distinguishing the cause of the accident but not the outcome; after E1's return and during the Questions phase, these children were also able to distinguish the harmful from non-harmful outcome conditions (likely with the aid of E1's sad emotional display in the harmful conditions).

Looks to E2. This ANOVA revealed a cause x outcome x age interaction, $F(1,118) = 5.43$, $p = .021$, $\eta_p^2 = .044$. Among 2-year-olds, there were no significant main effects or interactions, all ps > .245. Among 3-year-olds, there was a significant cause x outcome interaction, $F(1,60) = 9.31$, $p = .003$, $\eta_p^2 = .134$: Three-year-olds looked to E2 a significantly greater proportion of time if E2 had caused harm ($M = 6.97$, $SD = 5.53$) than if they had caused harm ($M = 1.59$, $SD = 1.71$; $t(30) = 3.72$, $p = .002$), whereas when the outcome was non-harmful, they looked to E2 a similar percentage of time regardless of who caused the outcome (Child-No Harm: $M = 2.02$, $SD = 1.76$; E2-No Harm: $M = 2.22$, $SD = 3.13$; $p = .831$). Notably, children's looking to E2 in E2-Harm was also significantly higher than in Child-No Harm and E2-No Harm (both ps < .007), whereas the other three conditions did not differ significantly from one another (all ps > .479). Thus, specifically when E2 had caused harm, 3-year-olds paid particular attention to E2, perhaps expecting her to respond in some way (e.g., accept responsibility, repair the damage). This supports our conclusion that 3-year-olds tracked both cause and outcome, and responded distinctly depending on who caused the outcome and whether or not the outcome was harmful.

Looks to tower. This analysis revealed a cause x outcome interaction, $F(1,118) = 3.99$, $p = .048$, $\eta_p^2 = .033$. Across ages, when harm was caused, children looked to the broken tower to a similar degree regardless of cause (Child-Harm: $M = 27.82$, $SD = 15.19$; E2-Harm: $M = 25.18$, $SD = 14.78$; $p = .487$), whereas when no harm had been caused, children looked to the broken tower for a greater proportion of time when they themselves had broken it ($M = 27.84$, $SD =$

12.26) than when E2 had broken it (M = 15.64, SD = 10.97; $t(61) = 4.16$, $p < .0005$). Notably, children's looking to the tower in E2-No Harm was also significantly lower than in Child-Harm and E2-Harm (both ps < .006), whereas the other three conditions did not differ significantly from one another (all ps > .439). These findings indicate that after E1 returned to the scene, children at both ages tracked both cause and outcome. Moreover, they were least concerned about the broken tower when they had not caused it to break and when it did not harm E1.

The ANOVA also revealed a significant cause x age interaction, $F(1,118) = 5.13$, $p = .025$, $\eta_p^2 = .042$. Two-year-olds looked significantly more to the broken tower if they had caused the outcome (M = 31.19, SD = 16.15) than if E2 had caused it (M = 18.34, SD = 10.17; $t(60) = 3.78$, $p < .0005$), whereas there was no significant difference between 3-year-olds' looking in the Child versus E2 conditions, $p = .534$.

Table S1

Coding scheme

Phase	Category (Scoring)	Description
E1's return	Attempt repair (1/0)	At least once during this phase, child attempts to repair picture tower by placing fallen blocks onto the part of the picture still standing. Stacking blocks to build a separate tower was not coded as repair.
	Guilt-related speech (1/0)	At least once during this phase, child's speech falls into one of the following categories: a) Apology b) Child offers to repair the picture (e.g., "I'll fix it," "We can repair it") c) Child states that he/she did not intend to cause the mishap or harm
Questions	Attempt repair (1/0)	At least once during this phase, child attempts to repair tower by placing fallen blocks onto the part of the picture still standing. Stacking blocks to build a separate tower was not coded as repair.
	Guilt-related speech (1/0)	At least once during this phase, child's speech falls into one of the following categories: a) Apology b) Child offers to repair the picture (e.g., "I'll fix it," "We can repair it") c) Child states that he/she did not intend to cause the mishap or harm
Individual questions		
Question 1 ("What happened?")	Statement about mishap (1/0)	Child makes at least one statement about the mishap (e.g., "The tower broke," "Broken")
Question 2 ("Who did it?") & Question 3 ("Did you do it?")	Accept blame (1/0)	At least once in response to either question, child either accepts responsibility for the mishap (e.g., "Yes" or nodding head in response to "Did you do it?") or blames him- or herself for the mishap (e.g., "I did it" in response to "Who did it?")
	Deny blame (1/0)	At least once in response to either question, child either denies responsibility for the mishap (e.g., "No" or shaking head in

		response to "Did you do it?") or blames someone or something else (e.g., "She [E2] did it," "It was a storm," or pointing to E2 in response to "Who did it?")
Question 4 ("What can be done about it?")	Statement about repair (1/0)	Child makes at least one statement about repair (e.g., "It can be repaired" or "You should repair")
Reassembling	Number of blocks given to E1 (raw number of blocks given)	Child gives a block to E1 either by extending her arm towards E1 or by placing it near her. If a child held out her arm with the clear intention of giving a block but E1 did not take it (e.g., because she did not notice it), it was included in the count. Blocks that the child picked up but placed farther away from E1 or with the intention of building a tower were not included.
Stickers	Number of stickers given to E1 (raw number of stickers given)	Child gives a sticker to E1 either by extending her arm towards E1 or by placing it near her. If a child held out her arm with the clear intention of giving a sticker but E1 did not take it (e.g., because she did not notice it), it was included in the count.

Note. The "Questions" phase refers to the entire phase, including all four questions. In addition, the individual questions were coded for specific responses and are thus also listed individually. The parentheses under each category indicate how the particular behavior was scored. Thus, "1/0" indicates that a score of 1 was assigned if the behavior was present in that phase, and a 0 was assigned if the behavior was absent in that phase.

Table S2

Number of children by age and condition who received each score for guilt-relevant reparative behavior (15 s following accident)

Condition		3-year-olds			2-year-olds		
	Score	0	1	2	0	1	2
Child-Harm		15	1	0	12	4	0
Child-No Harm		12	4	0	10	6	0
E2-Harm		16	0	0	15	0	0
E2-No Harm		16	0	0	14	1	0

Table S3

Number of children by age and condition who received each guilt-relevant reparative behavior score in the non-verbal ('Attempts to repair') and verbal ('Guilt in speech') categories (E1's return and Questions phases combined)

Condition		Attempts to repair						Guilt in speech					
		3-year-olds			2-year-olds			3-year-olds			2-year-olds		
	Score	0	1	2	0	1	2	0	1	2	0	1	2
Child-Harm		10	4	2	13	2	1	9	4	3	15	1	0
Child-No Harm		15	1	0	11	5	0	13	3	0	16	0	0
E2-Harm		16	0	0	14	0	1	15	1	0	15	0	0
E2-No Harm		15	1	0	14	1	0	16	0	0	15	0	0

Table S4

Mean number (and standard deviation) of blocks children gave to E1 in the Reassembling phase

Condition	3-year-olds	2-year-olds
Child-Harm	1.25 (1.69)	1.06 (2.54)
Child-No Harm	.56 (.89)	.44 (1.21)
E2-Harm	1.31 (1.62)	1.07 (2.28)
E2-No Harm	.38 (.72)	.60 (1.45)